Journal of Geochemical Exploration 139 (2014) 207–216

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/jgeoexp

Estimating the average concentration of minor and trace elements in surficial sediments using fractal methods

Tuhua Ma [a], Changjiang Li [a,*], Zhiming Lu [b]

[a] Zhejiang Information Center of Land and Resources, 310007 Hangzhou, China
[b] Computational Earth Science Group (EES-16), Los Alamos National Laboratory, Los Alamos, NM 87545, USA

ARTICLE INFO

Article history:
Received 29 January 2013
Accepted 5 August 2013
Available online 14 August 2013

Keywords:
Fractal distribution
Fractal averaging
Minor and trace elements
Censored data
Environmental geochemistry

ABSTRACT

The methods chosen to calculate the average value of the concentration for any geochemical element should depend on the probability distribution of the element abundance data. In this study, a fractal-based method was introduced to estimate the mean concentrations of geochemical elements that follow fractal frequency distributions. The fractal-based method has been tested on two abundance datasets for Ag, As, Au, Cu, Pb, Zn, Ce, Cr, and U from 529 floodplain sediment samples in China and from 10,927 stream sediment samples in Zhejiang Province, China. We compared the fractal method with other methods, including the arithmetic averaging, geometric averaging, and median, and found that there exist large discrepancies among these averages. The results show that the average calculated using the fractal-based method is always smaller than the arithmetic average and also generally smaller than the geometric mean and the median. The discrepancies may be attributed to the fact that the datasets follow a fractal distribution rather than a normal or a lognormal distribution. This study indicates that calculated arithmetic mean, geometric mean, or median may overestimate the average concentrations for elements that follow a fractal distribution.

1. Introduction

Estimating abundances of geochemical elements in the Earth's crust is always a challenge for geochemists, and it has attracted the attention of geochemists for at least 100 years. Element abundances are pivotal background values for exploration of mineral resources and determination of environmental pollution levels. In all existing models for estimating chemical composition of the crust (e.g., Clarke, 1889; Clarke and Washington, 1924; Gao et al., 1998; Goldschmidt, 1933; Taylor, 1964; Taylor and McLennan, 1985; Wedepohl, 1995), element abundances were derived from the averages of the compositions of surface exposures. The key difficulties in deriving element abundances include: (1) the tremendous geochemical heterogeneity of the crust, which calls for a need to devise a method for the generalization of particular data (Yaroshevskii, 2007); and (2) the reliability of the estimated mean concentration. Sediments from floodplains (Darnley et al., 1995; Xie and Cheng, 1997) and from continental river discharges (Yaroshevskii, 2007) have been considered as an "average sample" for materials of the crust exposed on erosion surfaces of continents, to overcome the first difficulty. However, the second problem still exists so far. Based on the assumption that concentrations in Earth's rocks and sediments follow normal distributions, average concentrations of geochemical elements were calculated by arithmetic averages (Rock, 1988). It has been found (Ahrens, 1954a,b) that many elements, especially trace elements, do not follow a normal distribution, but instead show a skewed or a tailed distribution. Calculating the arithmetic mean for skewed data will result in a biased (over) estimate of the central value (Filzmoser et al., 2009a). In this case, data are usually transformed by taking their logarithms, and then the normal model is used to calculate the geometric mean for the dataset. However, in many cases, a geometric mean is not appropriate because logarithms of data may still exhibit a heavy skewed or a tailed distribution (Chapman, 1976; Iqbal and John, 2010). Filzmoser et al. (2009a), based on the concept of compositional data analysis, suggested that when using an ilr-transformation for original data, it is possible to transform the computed arithmetic mean of ilr-transformed data back to the original data scale. However, as shown in Fig. 3 of Filzmoser et al. (2009a), the distribution of Na_2O is left-skewed; and it remains heavily left-skewed after ilr-transformation.

The heavy skewed or tailed distribution mentioned above, together with similar empirical discovery in many other application fields, has resulted in the formulation of the fractal theory (Mandelbrot, 1983). Original measured data from geochemical surveys can be regarded as a height field defined over a certain domain. The pattern of a height field is termed as geochemical landscape (geochemical surface). For many minor or trace elements, the geochemical landscapes, like the length of the famous Mandelbrot's coastline that commonly vary with the scale of observation (sampling density), are undetermined (Li et al., 2002, 2004), that is, in a given area, the denser one takes samples, the more details one can obtain. This implies that minor or

* Corresponding author.
E-mail addresses: zjigmr@mail.hz.zj.cn (C. Li), Zhiming@lanl.gov (Z. Lu).

0375-6742/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gexplo.2013.08.008

trace element abundance data probably follow fractal (power-law) distributions. In fact, many studies have provided strong support for the hypothesis that the distribution of minor or trace element abundance data is fractal (e.g., Allègre and Lewin, 1995; Bölviken et al., 1992; Cheng et al., 1994; Li et al., 2002, 2003, 2004; Lima et al., 2003).

If the abundance of a geochemical element follows a fractal distribution, the average value should be estimated with a fractal-based method rather than conventional methods such as arithmetic averaging or geometric averaging.

In this study, fractal averaging, a method based on fractal frequency distributions, is introduced to calculate the average concentration of geochemical elements. Hereinafter, the mean derived from the fractal averaging is called "fractal mean". This fractal-based method has been tested on two abundance datasets for Ag, As, Au, Cu, Pb, Zn, Ce, Cr, and U from 529 floodplain sediment samples that cover nearly most of the land surface of China and from 10,927 stream sediment samples collected in Zhejiang Province, China. We compared the fractal method with other methods, including the arithmetic averaging, geometric averaging, and median, and found that there exist large discrepancies among these averages.

2. Methodology

In general, a fractal (power-law) distribution is of the form

$$p(x) = Cx^{-\alpha}, \tag{1}$$

where $p(x)$ is the number of objects with size x, and C and α are constants, which can be determined from a dataset. The scaling exponent α could be a fraction and is usually called the fractal dimension. The density function diverges as x approaches zero, so there must be some low bound (denoted as $x_{\min}$) for this distribution. The normalization constant C can be found from the constraint $\int_{x_{\min}}^{\infty} p(x)\mathrm{d}x = 1$, i.e., $C = (\alpha - 1)x_{\min}^{\alpha-1}$, and the density function becomes

$$p(x) = \frac{\alpha-1}{x_{\min}}(x/x_{\min})^{-\alpha} \tag{2}$$

In many practical applications, one of the methods to study data is to calculate the cumulative distribution function (Newman, 2005). The probability $P(x)$ that the size X has a value greater than x is

$$P(x) = P(X>x) = \int_x^{\infty} p(x')\mathrm{d}x'. \tag{3}$$

If the distribution is fractal, substituting (2) into (3) and integrating it yields

$$P(x) = (x/x_{\min})^{-(\alpha-1)}. \tag{4}$$

Thus the cumulative distribution function $P(x)$ also follows a fractal distribution, but with a flatter slope (scaling exponent), which is 1 less than the exponent obtained from Eq. (1). The cumulative distribution has an advantage in that it can reduce statistical fluctuations without losing any information (Newman, 2005).

By definition, the ensemble mean for the fractal distribution can be derived as

$$\langle x\rangle = \int_{x_{\min}}^{\infty} xp(x)\mathrm{d}x = \frac{\alpha-1}{2-\alpha}x_{\min}\left(\frac{x}{x_{\min}}\right)^{-\alpha+2}\Bigg|_{x_{\min}}^{+\infty}. \tag{5}$$

It can be seen from Eq. (5) that the mean of the fractal distribution depends on the scaling exponent α, which leads to several scenarios. If

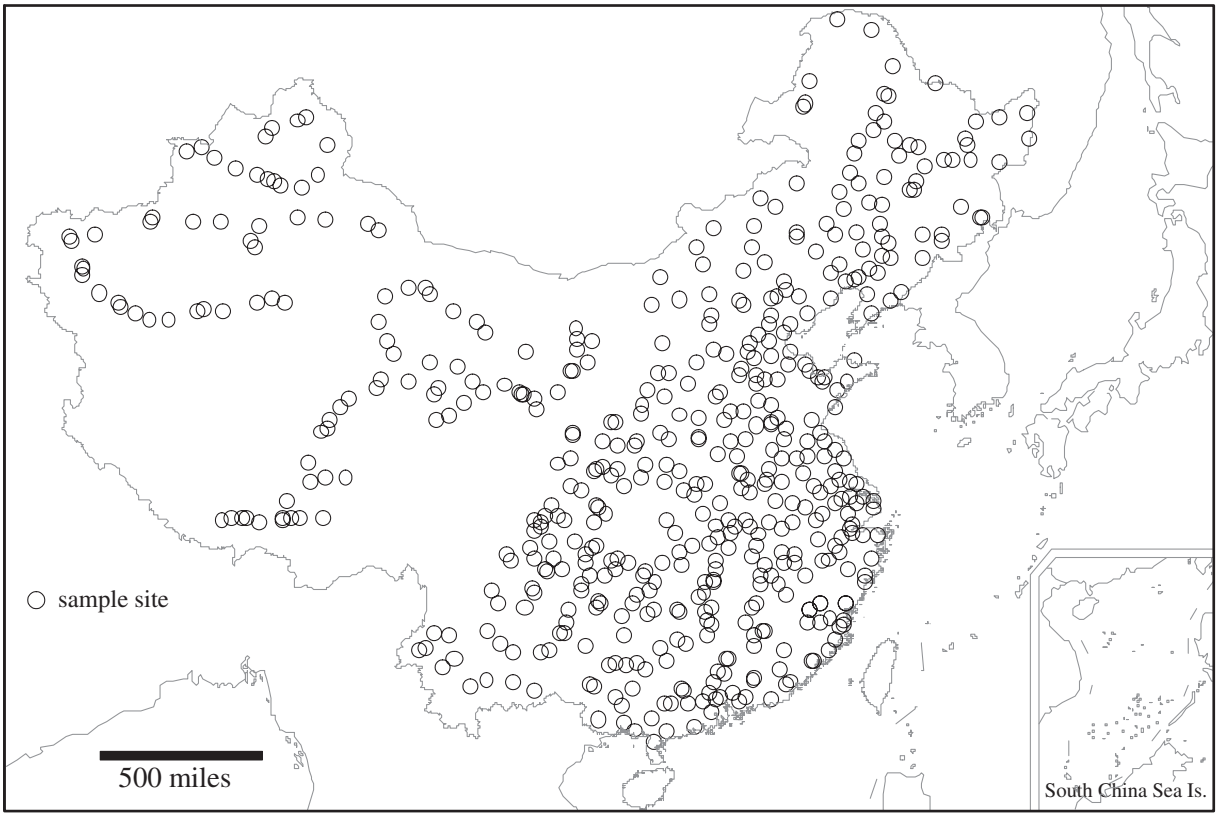

Fig. 1. Sampling locations of floodplain sediments for a wide-spaced geochemical survey in China.

$\alpha \le 2$, the fractal distribution does not have a finite mean. For $\alpha > 2$, the mean $\langle x \rangle$ given by Eq. (5) becomes

$$\langle x \rangle = \frac{\alpha-1}{\alpha-2} x_{\min}. \tag{6}$$

In this case, the mean square is given by

$$\left\langle x^2 \right\rangle = \int_{x_{\min}}^{\infty} x^2 p(x) \mathrm{d}x = \frac{\alpha-1}{3-\alpha} x_{\min}^2 \left(\frac{x}{x_{\min}} \right)^{-\alpha+3} \Bigg|_{x_{\min}}^{+\infty}. \tag{7}$$

This diverges if $\alpha \le 3$, which means that the fractal distribution has no finite mean square, and thus also no finite variance or standard deviation for $\alpha \le 3$. If $\alpha > 3$, then the mean square value can be given by

$$\left\langle x^2 \right\rangle = \frac{\alpha-1}{\alpha-3} x_{\min}^2. \tag{8}$$

In this case, the variance can be derived from the mean square minus the square of the mean:

$$\sigma^2 = \left\langle x^2 \right\rangle - \langle x \rangle^2 = \frac{\alpha-1}{(\alpha-3)(\alpha-2)^2} x_{\min}^2. \tag{9}$$

In many applications, the maximum of the distribution may be approximated using the largest sampled value (denoted as $x_{\max}$), and the mean and the mean square may be estimated as

$$\langle x \rangle = \frac{\alpha-1}{\alpha-2} \left[1 - \left(\frac{x_{\max}}{x_{\min}} \right)^{2-\alpha} \right] x_{\min} \tag{10}$$

and

$$\left\langle x^2 \right\rangle = \frac{\alpha-1}{\alpha-3} \left[1 - \left(\frac{x_{\max}}{x_{\min}} \right)^{3-\alpha} \right] x_{\min}^2 \tag{11}$$

and the variance can be calculated from $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$. It should be pointed out that both Eqs. (10) and (11) are valid for $\alpha > 1$.

3. Illustrative examples

In the International Geochemical Mapping Project (IGCP259), it was proposed to collect 5000 samples from each of all 160 × 160-km cells (25,600 km^2 in area) that cover the entire land surface of the world so as to make global geochemical maps and thus to obtain a better understanding of the elements' distributions over the continents (Darnley et al., 1995; Xie, 1996).

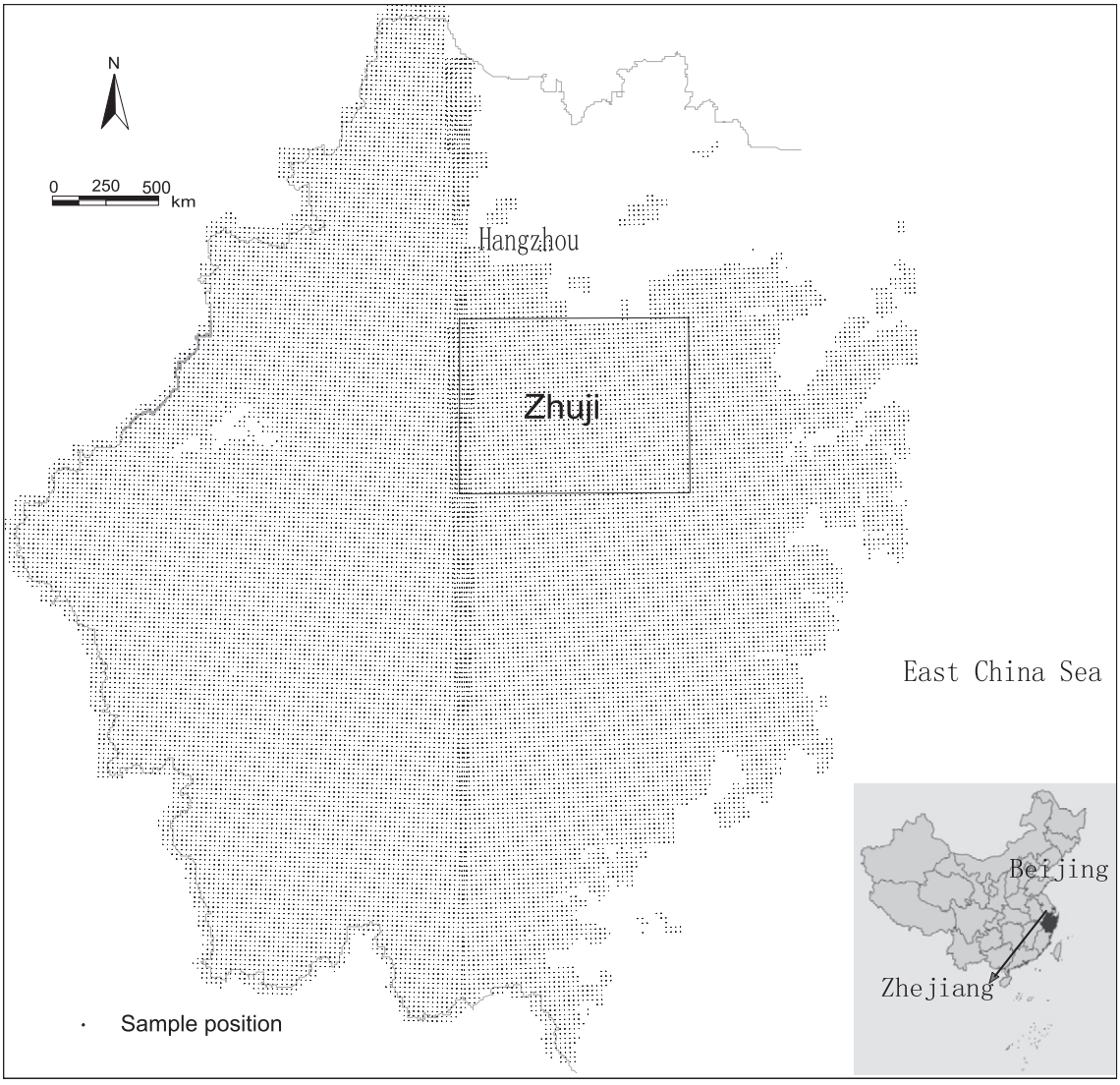

Fig. 2. Locations of stream sediment sampling stations in Zhejiang Province, China. The rectangular area in the figure shows the range of Zhuji City (see the text). Inset shows the location of Zhejiang in China.

Table 1
Comparison of the arithmetic mean, geometric mean, median, and fractal mean for some minor and trace elements in 529 samples from floodplain sediments, China.

Element	Detection method	Detection limit	Minimum reporting value	Maximum reporting value	Median	MAD	Arithmetic mean	Geometric mean	Fractal mean	Arithmetic variance	Fractal variance	$\alpha(=\beta+1)$	X_{min}	X_{max}
Ag	AAS	0.02 (ppm)	0.03	0.61	0.08	0.02	0.11	0.09	0.09	0.01	0.01	2.93	0.05	0.61
As	AFS	0.5 (ppm)	1.60	144	7.50	2.10	10.01	7.75	8.24	141.53	40.49	2.73	4.10	54
Au	AAS	0.2 (ppb)	0.20	52.50	1.50	0.60	2.06	1.56	1.58	8.53	0.95	2.90	0.88	7.20
Ce	XRF	3 (ppm)	10.0	214	70.00	14.00	74.13	69.42	64.85	747.63	399.53	4.93	49	214
Cr	XRF	7 (ppm)	15.00	334	70.00	12.00	73.48	68.61	68.94	1019.17	727.74	4.37	49	334
Cu	XRF	2 (ppm)	4.80	110.50	23.30	5.40	25.72	23.59	16.58	152.70	66.11	3.90	11	110.50
Pb	XRF	5 (ppm)	7.00	347	26.00	6.00	31.32	27.74	25.53	491.86	330.26	3.38	15	347
U	COL	1 (ppm)	1.30	12.50	1.90	0.30	2.19	2.06	1.97	1.10	0.94	3.86	1.3	12.50
Zn	XRF	10 (ppm)	3.00	618	75.00	18.00	80.40	71.54	48.43	2288.79	365.13	4.20	34	198

AAS: atomic absorption spectrometry; AFS: atomic fluorescence spectrometry; COL: colorimetry; XRF: X-ray fluorescence spectrometry.

In support of this effort, a project designed to test the feasibility of the wide-spaced global geochemical sampling and mapping, "Environmental Geochemistry Monitoring Networks and Dynamic Geochemical Maps in China", was launched in 1992. In 1993–1994, 529 sampling stations were established to carry out a wide-spaced floodplain sediment sampling, with an average density of one sample per 15,000 km^2, which irregularly covers most of the land surface of China (Fig. 1) (Cheng et al., 1997; Xie and Cheng, 1997). Two samples were collected at each sampling station: (a) surface floodplain sediment at a depth of 5–25 cm, and (b) deep floodplain sediment at a depth of 80 − 120 cm. The concentrations of 50 elements were measured for each sample. The detailed description of the sample medium and analytical methods of the floodplain sediment survey have been given in Cheng et al. (1997). On the other hand, as a sub-project of China's national geochemical mapping project (Regional Geochemistry–National Reconnaissance), 10,927 samples were collected from active stream sediments in Zhejiang Province (Fig. 2, about 101,800 km^2 in area) of China by the Zhejiang Geophysical and Geochemical Exploration Institute during 1980–1986. The sampling stations were set at the mouth of first-order streams or in the connected second-order streams. At each sampling station, sediment was gathered at four points with an average sampling density of 1 point per square kilometer. Samples from each station were composed of an equal weight of sediment from these four sampling points. The minimum weight of each sample is 2.5 kg. The composite samples from these stations, with an average density of one sample per 4 km^2, were submitted to the laboratory for chemical analysis. The concentrations of 39 elements were measured for the 10,927 stream sediment samples.

In this study, as an illustration of the fractal approach to calculate the average concentration of elements, we selected Ag, As, Au, Cu, Pb, Zn, Ce, Cr, and U content data from the floodplain sediments (529 samples at a depth of 5–25 cm) and from the stream sediments (10,927 samples). As a comparison, a subset of data consisting of 2026 samples in a small area about 7000 km^2 (i.e., Zhuji City of Zhejiang Province, see the rectangular area in Fig. 2) was randomly selected from the original stream sediment data. The selected nine elements can be divided into two groups (minor or trace elements) with different geochemical properties, i.e., thiophile elements (Ag, As, Au, Cu, Pb, and Zn) and oxyphile elements (Ce, Cr, and U). The analytical methods, detection limits of element determinations, and minimum and maximum reporting values, as well as the total number of samples measured are listed in Tables 1, 2 and 3.

4. Result and discussion

To estimate the average concentration of the minor and trace elements using the fractal-based method, we first examined the concentration–frequency relationship of the element data from the 529 floodplain sediment samples. Fig. 3 plots the cumulative number of samples exceeding a certain element concentration against the element concentration for elements Ag, As, Au, Cu, Pb, Zn, Ce, Cr, and U on a log–log scale. From this figure we can see that the assay data of As, Ce, Cr, Cu, Pb and Zn exhibit approximately straight lines over ranges of high concentrations and a flattening of the curves in the low concentration ranges from near zero concentrations to detection limits (see blue vertical lines in Fig. 3). The assay data of Ag and Au show a significant curvature in the low concentration ranges, almost a straight line relationship for intermediate concentrations, and a clear deviation from the straight line relationship in the high concentration ranges (at the right-hand side of the distributions). The cumulative graph of U, which is different from the other elements, shows that the assay data form an approximately straight line for the entire concentration range (see Fig. 3).

A dataset can be considered as samples from a particular population. If the scale range of samples is narrower than that of the population as a whole then the samples are truncated or censored, therefore biased (Pickering et al., 1995). Many studies (e.g., Barton and Zoback, 1992; Pickering et al., 1995) suggest that most of the samples from geological power law distributions are truncated or censored due to the resolution

Table 2
Comparison of the arithmetic mean, geometric mean, median, and fractal mean for some minor and trace elements in 10,927 samples from stream sediments in Zhejiang Province, China.

Element	Detection Method	Detection limit	Minimum reporting value	Maximum reporting value	Median	MAD	Arithmetic mean	Geometric mean	Fractal mean	Arithmetic variance	Fractal variance	$\alpha(=\beta+1)$	X_{min}	X_{max}
Ag	AAS	0.02 (ppm)	0.01	5	0.09	0.03	0.12	0.09	0.09	0.03	0.01	2.85	0.04	5
As	AFS	1 (ppm)	0.69	127	6.40	2.15	8.37	6.71	4.81	64.27	26.14	2.95	2.40	127
Au	AAS	0.3 (ppb)	0.01	2640	0.96	0.42	2.28	1.00	0.91	1032.83	5.08	2.44	0.30	104.33
Ce	XRF	3 (ppm)	6.00	410	93.50	16.80	100.74	95.06	66.65	1365.72	366.13	5.54	52	410
Cr	XRF	7 (ppm)	0.20	402	21.60	10.80	31.73	23.27	10.96	797.89	23.12	4.47	7.8	402
Cu	XRF	1 (ppm)	0.10	3534.50	10.80	4.10	15.75	11.34	8.25	2410.77	70.88	2.93	4.1	173.40
Pb	XRF	2 (ppm)	0.37	7640	32.40	7.50	40.14	33.54	32.10	8184.12	1570.01	3.13	17	8183.39
U	COL	1 (ppm)	0.70	32.50	3.50	0.70	3.67	3.47	2.79	2.18	1.29	4.57	2.01	32.50
Zn	XRF	4 (ppm)	13.80	5453.20	76.00	17.20	90.65	79.97	78.33	9297.41	2169.88	3.62	49	769

AAS: atomic absorption spectrometry; AFS: atomic fluorescence spectrometry; COL: colorimetry; XRF: X-ray fluorescence spectrometry.

Table 3
Comparison of the arithmetic mean, geometric mean, median, and fractal mean for some minor and trace elements in 2026 samples from stream sediments in Zhuji City of Zhejiang Province, China.

Element	Detection method	Detection limit	Minimum reporting value	Maximum reporting value	Median	MAD	Arithmetic mean	Geometric mean	Fractal mean	Arithmetic variance	Fractal variance	$\alpha\,(=\beta+1)$	X_{min}	X_{max}
Ag	AAS	0.02 (ppm)	0.01	3.20	0.08	0.02	0.10	0.08	0.08	0.01	0.01	3.02	0.04	3.20
As	AFS	1(ppm)	1.00	530	8.40	2.60	10.55	8.51	4.00	201.17	7.06	3.03	4.30	8.00
Au	AAS	0.3 (ppb)	0.10	602.90	0.98	0.40	2.31	1.09	1.08	287.43	1.68	2.46	0.42	15.28
Ce	XRF	3 (ppm)	7.75	371.40	85.62	9.93	90.02	87.42	74.09	612.32	311.16	6.25	60	371.40
Cr	XRF	7 (ppm)	0.20	341.10	25.10	10.50	35.75	26.39	20.81	1256.17	449.47	2.83	10	341.10
Cu	XRF	1 (ppm)	0.20	1210.60	16.80	6.30	24.17	17.21	15.52	1392.80	126.84	2.85	8.10	98.40
Pb	XRF	2 (ppm)	0.37	9526	29.09	6.41	36.59	28.22	19.16	43998.80	70.48	3.97	13	89.11
U	COL	1 (ppm)	0.72	26.45	3.06	0.58	3.20	3	2.78	1.86	1.21	4.61	2.01	26.45
Zn	XRF	4 (ppm)	33.12	1302.70	79.43	17.63	92.44	83.14	70.45	5396.08	614.57	4.86	51	198.40

AAS: atomic absorption spectrometry; AFS: atomic fluorescence spectrometry; COL: colorimetry; XRF: X-ray fluorescence spectrometry.

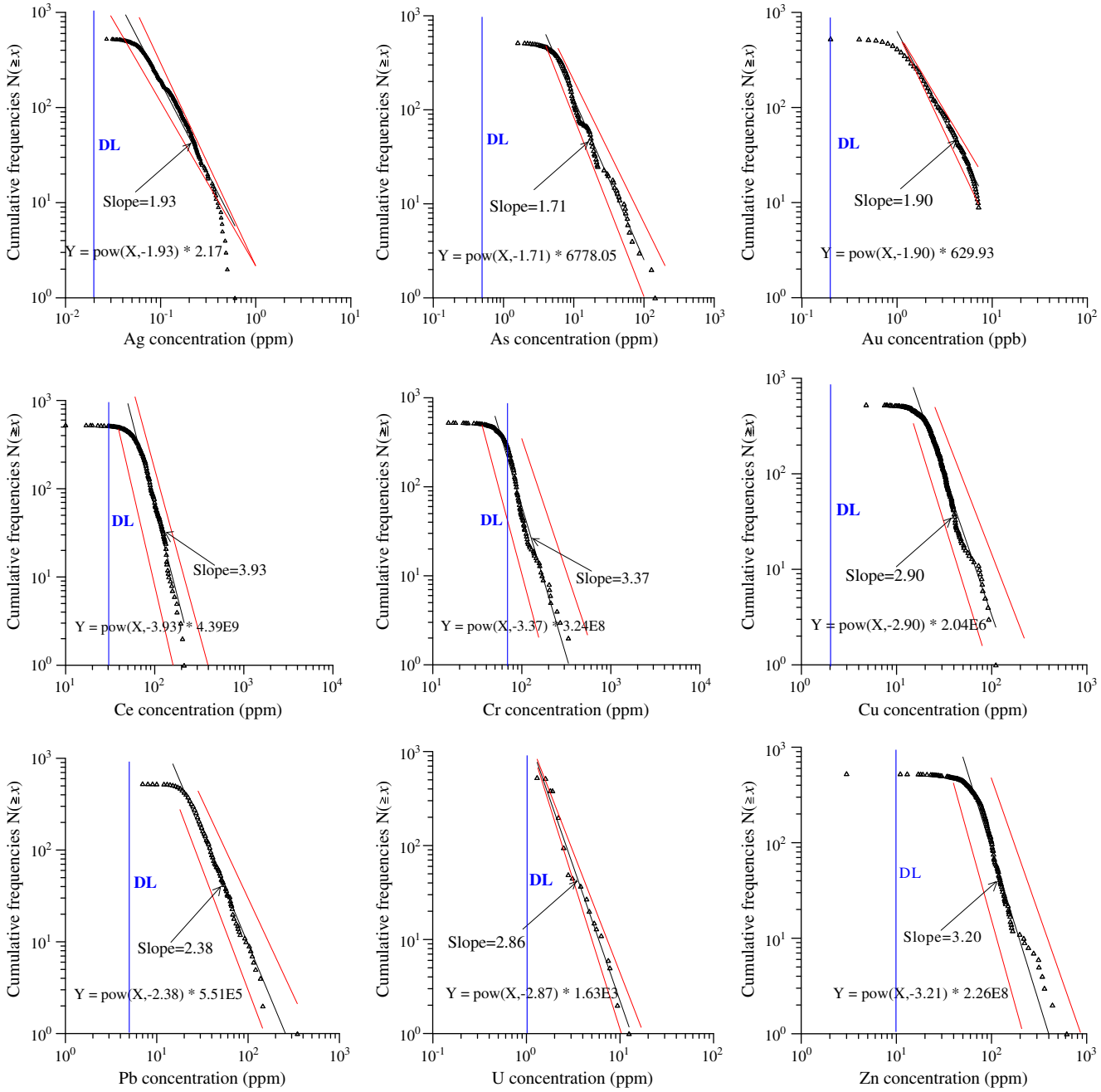

Fig. 3. Frequency distributions of Ag, As, Au, Ce, Cr, Cu, Pb, U, and Zn displayed in diagrams showing the logarithm of the cumulative number of samples exceeding a certain element concentration plotted against the logarithm of the element concentration. 529 samples from a floodplain sediment survey in China. DL (blue vertical lines) stands for the detection limit. The red line shows the 95% confidence limits [the standard deviations are $\pm 2\sigma$ based on Eq. (13)]. The slope of best-fit line (black line) is calculated using the least square fitting method.

limit of measurement techniques employed and the data recorded over a limited time (e.g., earthquake observation) or space (e.g., geological survey). Therefore, in Fig. 3, the left-hand fall-off and right-hand deviation observed in the log–log plots for concentration diagrams of Ag, As, Au, Cu, Pb, Zn, Ce, and Cr may be attributed to the following two effects.

First, the fall-off in the low concentration ranges may be related to the resolution of analytical methods. For example, many datasets in microanalysis or trace analysis, such as environmental testing, frequently contain values below the detection limit or limit of quantification of the analytical techniques due to the inherent limitations of chemical/analytical measurement methods (Zhao and Frey, 2004). These values are usually called left-censored data in environmental geochemistry (Antweiler and Taylor, 2008). The existing methods for treating left-censored data mainly include two alternatives: either substituting a constant for all observations below the detection limit or using random or evenly spaced numbers from a uniform distribution between zero and detection limit for the censored observations (Antweiler and Taylor, 2008; Clarke, 1998). Both of these alternatives result in a left-hand truncation in the low concentration range (Fig. 4).

Secondly, a finite-range effect (Pickering et al., 1995) may explain the deviation in the right-hand side of the log–log plots in Fig. 4. The finite-range effect can pose a right-censoring at high element concentrations (Fig. 4). During geochemical exploration, a survey area is usually divided into many individual cells, and one or several samples are collected in each cell, which implies that the geochemical data collected over a limited space will often miss rarely high values as indicators of rare geochemical processes, such as mineralization, consequently, resulting in a fall-off at the right-hand side of the plots (e.g., Ag and Au in Fig. 3).

Moreover, there are two types of measurement errors in measuring element concentration using an analytical method. Over a range of near zero concentrations, the measurement error is seen to be approximately constant; and over ranges of high concentrations, the measurement error is observed to be proportional to the concentration of an analyte (Currie, 1968; Hubaux and Vos, 1970). There are some difficulties in estimating the overall precision of an analytical method for data that span the "gray area" where a transition occurs between near zero concentrations and quantifiable amounts due to the existence of two types of measurement errors (Wilson et al., 2004). The fall-off at the left-hand side of the log–log plots in Fig. 4 can be considered to correspond with the gray area.

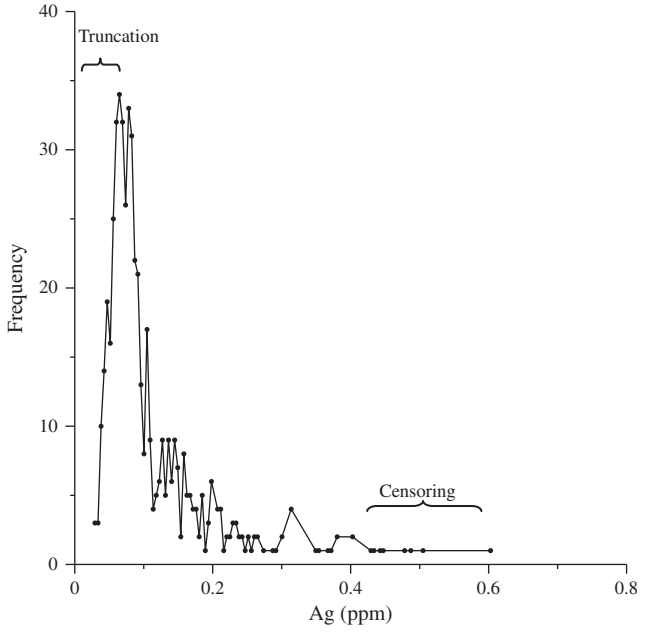

Fig. 4. Histogram displaying the frequency distribution of Ag assay data in floodplain sediment samples collected in China.

Any data within either the fall-off or censored data are not representative of the true distribution, and thus corrections are necessary (Barton and Zoback, 1992; Pickering et al., 1995). Pickering et al.(1995) developed an iterative approach to correct a cumulative graph with the left-hand fall-off and right-hand deviation in a log–log scale. The approach can be expressed as

$$\log N_C = \mathrm{Log}\, N_T - \beta_T(\log X_{MAX} - \log X_{MIN}) \tag{12}$$

where X_{MIN} and X_{MAX} are respectively the minimum and maximum values in a self-similar sub-sample data from a power-law population, N_C is the cumulative number of samples assigned to X_{MAX}, N_T is the cumulative number of samples assigned to X_{MIN}, and β_T is the slope of the straight line fitted to Eq. (4) for the sub-sample data taken from X_{MIN} to X_{MAX}.

For a cumulative graph fitted using a least squares fitting method, the standard deviation for the graph corrected by applying the iterative approach is given as (Pickering et al., 1995)

$$\beta \geq 1, \sigma = \frac{k\beta}{\sqrt{N}} \tag{13}$$

and

$$\beta < 1, \sigma = k\sqrt{\frac{\beta}{N}} \tag{14}$$

where σ is the standard deviation, k is a coefficient related to the sample size (magnitude) N, and $\beta = \alpha - 1$ is the slope of the straight line fitted to Eq. (4) using the least squares method. Based on Eq. (13) or (14), the 68% or 95% confidence limits for the estimated β values correspond to the mean plus/minus one or two standard deviations, respectively.

As shown in Fig. 3, the cumulative graphs of all elements except for U are required to be corrected. The corrected cumulative graphs (black line) are also shown in Fig. 3. All standard deviations in the β values for the corrected cumulative graphs are $\pm 2\sigma$ (see red line in Fig. 3) based on Eq.(13), the 95th percentile confidence limits, which means that 95% of the total samples for every element lie in the confidence intervals.

Similarly, Fig. 5 shows the cumulative graphs for the elements Ag, As, Au, Cu, Pb, Zn, Ce, Cr, and U from the stream sediment samples collected in the Zhejiang region. Fig. 6 depicts the cumulative graphs for the same elements from the stream sediment samples in Zhuji City of Zhejiang Province. The cumulative graphs of Au, Cu, Pb, and Zn in Fig. 5, as well as Au, Pb, Zn (and possibly Ag) in Fig. 6, which are different from the other elements, show an upward deviation from the straight-line relationships at high element concentrations (near the right-hand side of plots). The data within this upward deviation may be related to locally high geochemical anomaly distributions, as indicated from field geological observations that some stream sediment samples were collected from areas near gold, silver, lead, zinc, or copper deposits in the Zhejiang region.

Following the same methodology used for Fig. 3, the cumulative graphs in Figs. 5 and 6 have been corrected (black line) and all standard deviations in the β values for the corrected cumulative graphs are $\pm 2\sigma$ (red line) based on Eq.(13), the 95th percentile confidence limits.

The scaling exponent β for a cumulative fractal (power-law) distribution, as shown in Section 2, is related to the scaling exponent α for the corresponding probability density functions (non-cumulative distribution): $\alpha = \beta + 1$. As illustrated in Figs. 3, 5, 6 and Tables 1, 2 and 3, the corrected cumulative graphs for every element concentration dataset can satisfy the conditions required by Eqs. (6), (7) and (8), i.e., $\alpha > 2$ and $\alpha \neq 3$.

Next, we calculate the fractal mean and variance of concentrations for the element data. In calculating the fractal mean, we take the minimum concentration level of the corrected cumulative graph (X_{MIN}) as

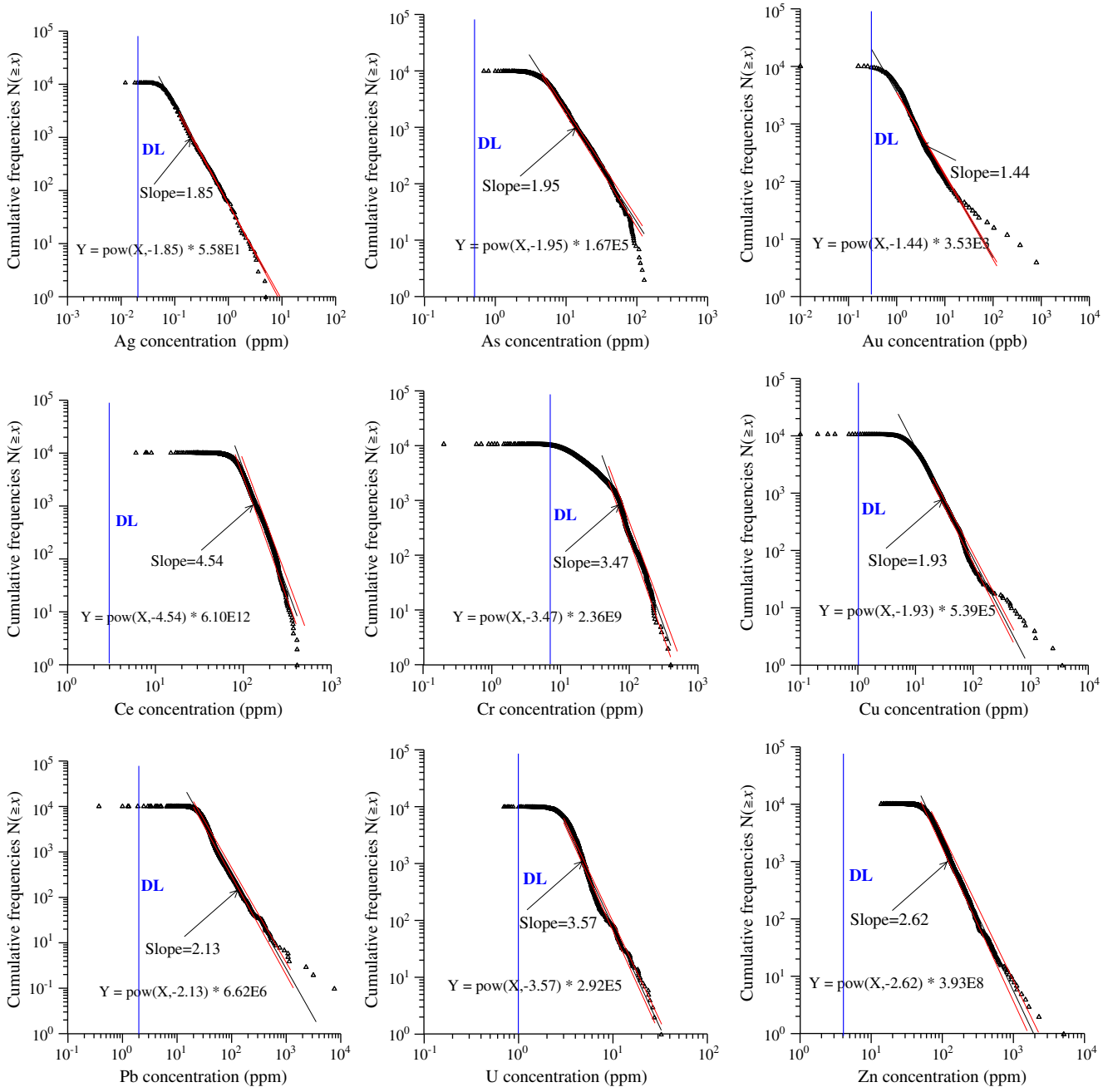

Fig. 5. Frequency distributions of Ag, As, Au, Ce, Cr, Cu, Pb, U, and Zn showing the logarithm of the cumulative number of samples exceeding a certain element concentration plotted against the logarithm of the element concentration. 10,927 samples from a stream sediment survey in Zhejiang Province. DL (blue vertical lines) stands for the detection limit. The red line shows the 95% confidence limits [the standard deviations are $\pm 2\sigma$ based on Eq. (13)]. The slope of best-fit line (black line) is calculated using the least squares fitting method.

x_{min} in Eq. (10). The variance is obtained from the mean square minus the square of the mean. For the calculation of the mean squares, the minimum and maximum concentration levels of the corrected cumulative graph (X_{MIN} and X_{MAX}) are taken as x_{min} and x_{max} in Eq. (11), respectively. The mean values and variances calculated using the fractal-based method as well as median and median absolute deviation (MAD) for Ag, As, Au, Cu, Pb, Zn, Ce, Cr, and U concentrations in the floodplain sediment samples and stream sediment samples are shown in Tables 1, 2 and 3.

Comparing the fractal mean with the commonly used arithmetic mean or geometric mean for the element concentration data (Fig. 7 and Tables 1, 2 and 3), as expected, large discrepancies exist among the three averages. For example, all fractal means are lower than the arithmetic means. The differences between the fractal means and the arithmetic means are between 6.6% (for Cr data in China) and 189.5% (for Cr data in Zhejiang). In particular, for As in Zhuji City and Cr in the Zhejiang region, the fractal means are respectively 1.6 and 1.9 times lower than their corresponding arithmetic means. The fractal means, except for the means of As, Au and Cr data in China, are also lower than their corresponding geometric means.

All medians are lower than the arithmetic means, but they can be either higher or slightly lower than the fractal means (see Fig. 7 and Tables 1, 2 and 3), which indicates that medians and fractal means are not compatible. Such incompatibility may stem from the difference on

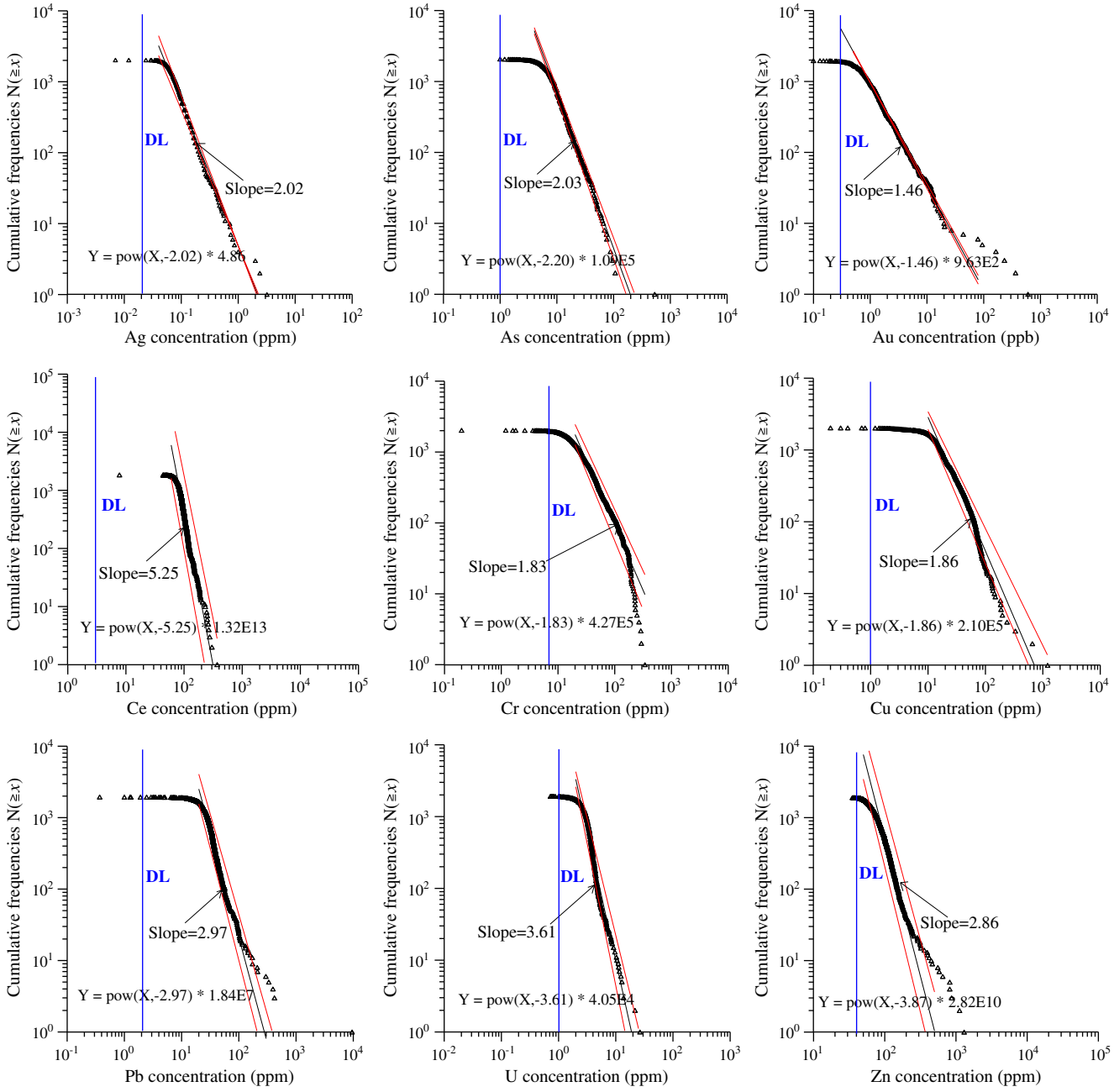

Fig. 6. Frequency distributions of Ag, As, Au, Ce, Cr, Cu, Pb, U, and Zn showing the logarithm of the cumulative number of samples exceeding a certain element concentration plotted against the logarithm of the element concentration. 2026 samples from a stream sediment survey in Zhuji city. DL (blue vertical) shows the detection limit. The red line shows the 95% confidence limits [the standard deviations are $\pm 2\sigma$ based on Eq. (13)]. The slope of best-fit line (black line) is calculated using the least squares fitting method.

the actual subset of data used in calculating medians and fractal means. The fractal means are derived from the corrected data (by ignoring the fall-off or censored data), while the medians are determined from the original samples. It is noted that the medians are sensitive to the number of data points in the sample, although they are insensitive to the actual values of outliers. This also explains the incompatibility between the median absolute deviation (MAD) and the fractal variances.

The different averages are obtained from the same data set using different methods. The question is "which of these averages is the most credible?" This question cannot be answered at this point, because the true mean value of any element concentration is unknown. However, we should emphasize that the choice of methods used for calculating the average value of concentration for geochemical elements depends on the distribution pattern of datasets. If the distributions of element concentration (especially minor or trace elements) in Earth's rocks and sediments tend to a fractal pattern, their mean values should be estimated with the method based on the fractal concept. The variance, or its square root, characterizes the concentrated extent of the data around the central value (mean value). As those shown in Tables 1, 2 and 3, all fractal variances are smaller than the arithmetic variances, which indicates that the fractal mean may be more appropriate than the arithmetic mean for the data.

It should be pointed out that geochemical data are compositional data (Aitchison, 1986; Filzmoser et al., 2009a,b; Reimann et al., 2012). There is a constraint on compositional data, i.e., the sum of all components being one (or 100%), which means that compositional data are

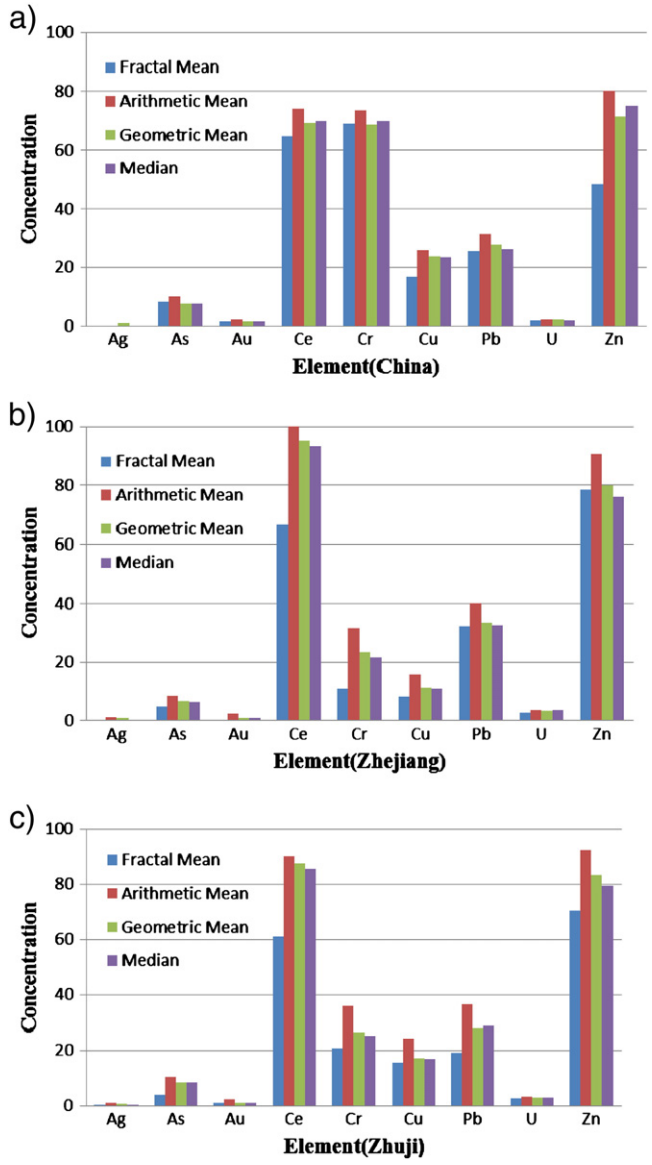

Fig. 7. Comparison of four methods for calculating mean of some minor and trace element data in surficial sediments: (a) China; (b) Zhejiang Province; and (c) Zhuji City.

closed. It has been suggested in the literature that some types of data transformation may be needed prior to any analysis (Aitchison, 1986; Filzmoser et al., 2009a,b). Several transformations have been proposed, including the additive logratio (alr), centered logratio (clr), and isometric logratio (ilr). For the purpose of comparison of the fractal mean and fractal variance derived from the data without transformation, we first made the centered logratio transformation, which can be expressed as (Aitchison, 1986; Aitchison and Egozcue, 2005)

$$clr(x) = [\ln(x_1/g(x)),\ \ln(x_2/g(x)), \ldots,\ \ln(x_D/g(x))] \quad (15)$$

where $g(x)$ is the geometric mean of the components of x defined by Aitchison (1986). After using the clr-transformation for the original data of Ag, As, Au, Cu, Pb, Zn, Ce, Cr, and U from the 529 floodplain sediment samples, we then plotted the cumulative number of samples exceeding a certain ratio $(x_i/g(x))$ plotted against the $(x_i/g(x))$ on a log–log paper. It is found that the shape of the log–log plots (not shown) is basically the same with that of Fig. 3, which indicates that if geochemical element abundance data follow a fractal distribution, the logratio-transformed data still remain the fractal distribution.

We noted that, in a number-average size model which shows a power-law (fractal) relationship with the average size, the average size is defined as the mean of the sizes of the objects beyond a given size (Wang et al., 2011). However, as shown in Eq. 9 of Wang et al. (2011), the mean is an arithmetic average calculated using a moving average method, which is not based on the fractal concept.

Finally, it is interesting to point out that the left-censored data in environmental measurements were commonly evaluated by assuming that data follow a normal or lognormal distribution (Antweiler and Taylor, 2008). However, if concentration data of trace elements in environmental testing follow a fractal (power-law) distribution, the fractal approach may be more meaningful than the traditional methods for evaluating the left-censored data and simulating the "true values" hidden in the censored data.

5. Conclusions

Because minor or trace element abundance data in Earth's rocks and sediments often tend to follow a fractal frequency distribution, their mean values should be determined with the fractal-based method. In this study, we use a fractal method to estimate the average values of concentration for geochemical elements. The key points in this method are first to plot the cumulative number of samples exceeding a certain element concentration plotted against the element concentration on a log–log paper. If the cumulative number of samples as function the element concentration falls quite closely on a straight line with slope $\beta > 1$ and $\beta \neq 2$, then one should apply the derived fractal frequency relation to calculate the mean value of the element abundance data. However, cumulative graphs for most datasets from geochemical surveys are biased (truncated or censored) due to the resolution limit of chemical/analytical measurement methods and over a limited space. Thus the cumulative graphs need to be corrected, and the fractal means can be calculated using these corrected graphs.

We illustrated the method using two abundance datasets for elements Ag, As, Au, Cu, Pb, Zn, Ce, Cr, and U, which have different geochemical properties, from 529 floodplain sediment samples and from 10,927 stream sediment samples. These examples demonstrate that the differences among the fractal means, arithmetic means, geometric means, and medians for the same datasets are very large. All the fractal means are lower than the arithmetic means and also generally lower than the geometric means and the medians. This is because the data appear more likely a fractal rather than a normal or a lognormal distribution.

The evaluations of assay data falling between near zero concentrations and quantifiable amounts, especially the left-censored data, are still critical issues in environmental measurements. This study showed that, in many cases, fractal (power-law) statistics may be of considerable value for evaluating the left-censored data.

Acknowledgments

We thank Prof. Xuejing Xie of the Institute of Geophysical & Geochemical Exploration, Chinese Academy of Geological Sciences for his sharing 529 wide-spaced geochemical sampling data. We would like to thank the editor and two anonymous reviewers for their valuable comments and suggestions, which have improved the paper.

References

Ahrens, L.H., 1954a. The lognormal distribution of the elements (a fundamental law of geochemistry and its subsidiary). Geochimica et Cosmochimica Acta 5, 49–73.
Ahrens, L.H., 1954b. The lognormal distribution of the elements—II. Geochimica et Cosmochimica Acta 6, 121–131.
Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Chapman & Hall, London (416 pp.).
Aitchison, J., Egozcue, J.J., 2005. Compositional data analysis: where are we and where should we be heading? Mathematical Geology 37 (7), 829–850.
Allègre, C.J., Lewin, E., 1995. Scaling laws and geochemical distribution. Earth and Planetary Science Letters 132, 1–13.

Antweiler, R.C., Taylor, H.E., 2008. Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. Environmental Science and Technology 42, 3732–3738.
Barton, C.A., Zoback, M.D., 1992. Self similar distribution and properties of macroscopic fractures at depth in crystalline rock in the Cajon Pass Scientific borehole. Journal of Geophysical Research 97, 5181–5200.
Bölviken, B., Stokke, P.R., Feder, J., Jössang, T., 1992. The fractal nature of geochemical landscapes. Journal of Geochemical Exploration 43, 91–109.
Chapman, R.P., 1976. Some consequences of applying lognormal theory to pseudolognormal distributions. Mathematical Geology 8, 209–214.
Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. Journal of Geochemical Exploration 51, 109–130.
Cheng, H.X., Shen, X.C., Yan, G.S., Gu, T.X., Lai, Z.M., Xie, X.J., 1997. Wide-spaced floodplain sediment sampling covering the whole of China: pilot survey for International Geochemical Mapping. In: Xie, X.J. (Ed.), Proceedings of the 30th International Geological Congress. 19, Geochemistry, pp. 89–109.
Clarke, F.W., 1889. The relative abundance of the chemical elements. Philosophical Society of Washington, Bulletin 11, 131–142.
Clarke, J.U., 1998. Evaluation of censored data methods to allow statistical comparisons among very small samples with below detection limit observations. Environmental Science Technology 32 (1), 177–183.
Clarke, F.W., Washington, H.S., 1924. The composition of the Earth's crust. USGS Professional Paper 127, 117pp.
Currie, L., 1968. Limits for qualitative detection and qualitative determination—application to radiochemistry. Analytical Chemistry 40, 587–593.
Darnley, A.G., Björklund, A., B¨olviken, B., Gustavsson, N., Koval, P.V., Plant, P.A., Steenfelt, A., Tauchid, M., Xie, X.J., 1995. A global geochemical database for environmental and resource management. Recommendations for International Geochemical Mapping: Final Report of IGCP Project 259. Earth Sciences 19, 41–45 (UNESCO Publishing).
Filzmoser, P., Hron, K., Reimann, C., 2009a. Univariate statistical analysis of environmental (compositional) data: problems and possibilities. Science of the Total Environment 407, 6100–6108.
Filzmoser, P., Hron, K., Reimann, C., 2009b. Principal component analysis for compositional data with outliers. Environmetrics 20 (6), 621–632.
Gao, S., Luo, T.C., Zhang, B.R., Zhang, H.F., Han, Y.W., Hu, Y.K., Zhao, Z.D., 1998. Chemical composition of the continental crust as revealed by studies in east China. Geochimica et Cosmochimica Acta 62, 1959–1975.
Goldschmidt, V.M., 1933. Grundlagen der quantitativen Geochemie. Fortschrift Mineralogie 17 (2), 112–156.
Hubaux, A., Vos, G., 1970. Decision and detection for linear calibration curves. Analytical Chemistry 42, 849–855.
Iqbal., M.F., John, L.K., 2010. Confusion by all means. Proceedings of the 6th International Workshop on Unique Chips and Systems (UCAS-6).
Li, C.J., Zhao, N.L., Ma, T.H., 2002. Fractal reconstruction with unorganized geochemical data. Mathematical Geology 34 (7), 809–829.
Li, C.J., Ma, T.H., Shi, J.F., 2003. Application of a fractal method relating concentrations and distances for separation of geochemical anomalies from background. Journal of Geochemical Exploration 77, 167–175.
Li, C.J., Ma, T.H., Chen, J.J., 2004. A fractal interpolatory approach to geochemical exploration data processing. Mathematical Geology 36 (5), 593–606.
Lima, A., De Vivo, B., Cicchella, D., Cortini, M., Albanese, S., 2003. Multifractal IDW interpolation and fractal filtering method in environmental studies: an application on regional stream sediments of Campania region (Italy). Applied Geochemistry 18, 1853–1865.
Mandelbrot, B.B., 1983. The Fractal Geometry of Nature. Freeman, New York.
Newman, M.E.J., 2005. Power laws, Pareto distributions and Zipf's law. Contemporary Physics 46, 323–351.
Pickering, G., Bull, J.M., Sanderson, D.J., 1995. Sampling power-law distributions. Tectonophysics 248, 1–20.
Reimann, C., Filzmoser, P., Fabian, K., Hron, K., Birke, M., Demetriades, A., Dinelli, E., Ladenberger, A., The GEMAS Project Team, 2012. The concept of compositional data analysis in practice — total major element concentrations in agricultural and grazing land soils of Europe. Science of the Total Environment 426, 196–210.
Rock, N.M.S., 1988. Summary statistics in geochemistry: a study of the performance of robust estimates. Mathematical Geology 20 (3), 243–275.
Taylor, S.R., 1964. Abundance of chemical elements in the continental crust—a new table. Geochimica et Cosmochimica Acta 28, 1273–1285.
Taylor, S.R., McLennan, S.M., 1985. The Continental Crust: Its Composition and Evolution. Blackwell, Oxford.
Wang, Q.F., Wan, L., Zhang, Y., Zhao, J., Liu, H., 2011. Number-average size model for geological systems and its application in economic geology. Nonlinear Processes in Geophysics 18, 447–454.
Wedepohl, H., 1995. The composition of the continental crust. Geochimica et Cosmochimica Acta 59, 1217–1239.
Wilson, M.D., Rocke, D.M., Durbin, B., Kahn, H.D., 2004. Detection limits and goodness-of-fit measures for the two-component model of chemical analytical error. Analytica Chimica Acta 509, 197–208.
Xie, X.J., 1996. Exploration geochemistry: present status and prospects. Geological Review 42, 346–356 (in Chinese with English abstract).
Xie, X.J., Cheng, H.X., 1997. The suitability of floodplain sediment as global sampling medium. Evidence from China. Journal of Geochemical Exploration 58, 51–62.
Yaroshevskii, A.A., 2007. On the reliability of estimates of the average chemical composition of the upper continental crust and river discharge. Geochemistry International 45 (11), 1144–1145.
Zhao, Y., Frey, H.C., 2004. Quantification of variability and uncertainty for censored data sets and application to air toxic emission factors. Risk Analysis 24 (3), 1019–1034.